

## Calíope: herramienta para gestionar un corpus y un glosario de términos informáticos

M<sup>a</sup> del Socorro Bernardos Galindo

Guadalupe Aguado de Cea

Facultad de Informática, Universidad Politécnica de Madrid.

### Resumen

En esta comunicación se presenta una aplicación, Calíope (Barahona 2006), diseñada con el objetivo de proporcionar a alumnos que estudian informática no sólo una forma de aprender el uso de los términos en contexto, sino también de ver las relaciones sintácticas y léxico-semánticas que se establecen entre ellos. Siguiendo los enfoques actuales en terminología (Cabré 2003, Temmerman, 2000) se ha creado esta aplicación en donde, además de una selección de los textos, atendiendo a los criterios más aceptados en el campo de la lingüística de corpus para las lenguas de especialidad (Bowker & Pearson 2002) se recopilan los términos dentro del contexto en el que aparecen, se obtienen las concordancias y las relaciones existentes entre ellos (Aguado 2007). Con esta finalidad, se ha construido una herramienta capaz de manejar on-line dos recursos: un corpus y un glosario de términos en inglés y en español. Este diseño resulta muy útil, por ejemplo, para traducir textos informáticos del inglés al español y viceversa, pues permite saber en qué contextos se utiliza cada término. En consonancia con esta característica su uso resulta también muy atractivo en el ámbito de la enseñanza de la lengua inglesa, o en la redacción de textos informáticos en inglés.

**Palabras clave:** lingüística de corpus, terminología informática, fraseología especializada

### Introducción

Hoy en día los trabajos en lingüística aplicada, sea con fines didácticos o investigadores no se conciben sin la utilización de un corpus (Sánchez 2000), que refleje el uso de la lengua en contexto. Cuando se pretende mostrar este uso a hablantes de una segunda lengua (L2) con un nivel intermedio de conocimiento de la L2 y con necesidades no sólo lingüísticas, sino también de conocimientos del dominio, la utilización de un corpus de textos técnicos se ha mostrado fundamental. Sin embargo, son pocos los corpus científico-técnicos bilingües<sup>1</sup> que cumplen los requisitos necesarios para satisfacer los objetivos de un determinado grupo dentro del ámbito de las lenguas de especialidad. Por otra parte, no es fácil encontrar software educativo que facilite la integración de tareas diversas de interés para los alumnos, desde las enfocadas al aprendizaje de gramática o el uso de la lengua, que son las más frecuentes, a las que aporten glosarios bilingües con términos extraídos de los contextos académicos, científicos o técnicos, teniendo en cuenta la diversidad de géneros con que se pueden encontrar los alumnos. Conviene señalar también que en la construcción de software intervienen otros factores que impiden la utilización inmediata de otros productos

---

<sup>1</sup> Entre los corpus multilingües de carácter científico-técnico cabe mencionar el del Instituto Universitario de Lingüística Aplicada (IULA) de la Universidad Pompeu Fabra, y su herramienta de explotación Bwananet. De carácter monolingüe, se encuentra el Corpus de Referencia del Español Actual (CREA), para la lengua española. En cambio, son muy numerosos los corpus elaborados para la lengua inglesa, (*British National Corpus* (BNC), entre otros, que generalmente no contemplan los tipos de textos científico-técnicos del ámbito informático y de Internet, al menos por ahora. Mención aparte merece la herramienta WebCorp, para búsqueda *on-line*, <http://www.webcorp.org.uk/>, la cual cuenta también con ciertas prestaciones que pueden ayudar a centrar las consultas.

anteriores: lenguaje de programación utilizado, entorno sobre el que se ha construido, interfaces utilizadas, etc. De ahí la necesidad de diseñar nuevas aplicaciones que, aunque globalmente pueden compartir objetivos docentes con otras ya en el mercado, resulten más adecuadas para los fines perseguidos. Teniendo esto en cuenta y siguiendo los enfoques actuales en terminología, (Cabré 2003, Temmerman, 2000, L'Homme 2000) que priorizan un enfoque lingüístico-textual de la terminología junto con su proyección cognitiva, se ha creado una aplicación, Calíope (Barahona 2006), en donde, además de una selección de los textos, atendiendo a los criterios más aceptados en el campo de la lingüística de corpus para las lenguas de especialidad (Bowker & Pearson 2002), se recopilan los términos dentro del contexto en el que aparecen, se obtienen las concordancias y se relacionan los conceptos propios de ese dominio con las estructuras léxico-sintácticas en las que se insertan los términos (Aguado 2007). Con esta finalidad se ha construido la herramienta que aquí se presenta, la cual es capaz de manejar *on-line* dos recursos: un corpus y un glosario de términos tanto en inglés como en español. El objetivo era doble: (a) proporcionar a alumnos que estudian informática (en concreto, en la Facultad de Informática de la Universidad Politécnica de Madrid) una forma de aprender el uso de los términos (en inglés) en contexto, y de ver los equivalentes habituales en lengua española, y (b) mostrar las relaciones sintácticas y léxico-semánticas que se establecen entre ellos tanto en una lengua como en la otra, con el fin de familiarizar al alumno con la fraseología especializada. En este sentido, se ha tratado de aplicar la metodología propuesta por L'Homme<sup>2</sup> para la lengua francesa.

En la siguiente sección se describe brevemente el corpus extraído como base de trabajo y el glosario que hasta este momento se ha recopilado, sin que puedan considerarse, en modo alguno, exhaustivos. A continuación se presentan, de forma somera, las principales características de esta herramienta y después se muestra cómo se puede emplear con fines didácticos. Finalmente se señalan las conclusiones a las que se ha llegado y las mejoras previstas.

### **Método de elaboración del corpus y glosario**

Tras un análisis de los tipos de textos que los alumnos han de manejar tanto en sus tareas discentes (preparación de prácticas, instalación de programas, construcción de aplicaciones, etc.) como pre-profesionales, se elaboró una lista de la tipología textual que podía ser relevante: artículos, manuales académicos y de instalación de dispositivos, tanto *hardware* como *software*, catálogos de productos informáticos, trabajos fin de carrera, etc. Hasta el momento, el corpus no está equilibrado en las dos lenguas, ya que se cuenta con 4.500.000 de palabras en inglés mientras que en español sólo se ha llegado a un millón de palabras. La razón de este desequilibrio radica en que la finalidad inicial era su aplicación al aprendizaje de inglés específico. Sin embargo, posteriormente, se ha observado la necesidad de contar también con textos en español, dadas las posibles necesidades de los nuevos alumnos Erasmus que vienen a nuestro centro.

En lo concerniente al glosario, a diferencia de los glosarios más tradicionales que centran la búsqueda en los términos de carácter nominal, en una primera fase se partió de los elementos verbales que generalmente acompañan a los nominales y que, según

---

<sup>2</sup> Esta autora ha desarrollado un diccionario de francés de informática e Internet, en lengua francesa, en el que se recogen estas relaciones siguiendo el modelo de la Teoría Sentido-Texto propuesto por Mel'cuk (1988) y sus colaboradores. <http://olst.ling.umontreal.ca/dicoinfo/>

nuestra experiencia plantean una gran dificultad en la redacción de textos en la L2. De esta manera, se pretendía ver las relaciones que se establecen entre elementos verbales y nominales, la fraseología específica y las posibilidades combinatorias que permiten, al tiempo que se han agrupado estas relaciones por conjuntos léxico-semánticos. Por ejemplo, se ha observado que *download* va seguido de términos que pertenecen a algún tipo de *software*: *file*, *data*, *program*.

### Características técnicas de Calíope

Entre las principales características de la aplicación cabe mencionar el hecho de que la interfaz está formada por páginas web, con lo que la navegación a través de estas páginas resulta fácil para el usuario, acostumbrado generalmente a navegar por Internet, es decir, a moverse a través de botones, menús y enlaces. Por otra parte, la aplicación puede utilizarse *on-line*, siempre que el usuario esté registrado y tenga el permiso adecuado. El sistema comprueba los datos del usuario y si son válidos despliega la ventana principal de la aplicación con los servicios disponibles para el tipo de usuario del que se trate: Los administradores del sistema podrán acceder a todos los servicios (administrar usuarios, administrar textos, consultar el corpus y administrar y consultar el glosario), pero los usuarios normales sólo podrán hacer consultas (del corpus y del glosario). A continuación se explica en qué consiste cada uno de estos servicios.

- Administrar usuarios. Un administrador puede dar de alta un nuevo usuario o un conjunto de usuarios (a partir de un archivo), eliminar un usuario y modificar los datos de un usuario.
- Administrar textos. Para cada texto se almacena información relativa a su tipo (por ejemplo, texto científico) y su fuente (por ejemplo, [www.ia.com](http://www.ia.com)), así como el campo de estudio al que pertenece (por ejemplo, inteligencia artificial) y su lengua. Además de dar de alta, eliminar o modificar los textos, un usuario administrador también puede dar de alta, eliminar o modificar los tipos, los campos y las fuentes de los textos. Para que la herramienta pueda manejar los textos, su formato ha de ser texto plano o HTML.
- Consultar el corpus. El sistema incorpora una serie de operaciones que permiten examinar el corpus y que son similares a los de otros concordancieros: listado y frecuencia de palabras, búsqueda de las concordancias de un término según varios parámetros, visualización del texto, etc. Los resultados pueden descargarse en un fichero HTML.
- Administrar el glosario. El glosario almacena términos y las relaciones existentes entre ellos. Un administrador puede dar de alta y eliminar un término y modificar información sobre él. Para cada término se almacena su lengua, categoría gramatical, definiciones de sus acepciones, traducción al otro idioma, número de ocurrencias en el corpus, contextos de aparición y relaciones semánticas con otros términos del glosario (por ejemplo, agente). Además, el administrador también puede acceder a las funciones necesarias para la administración de las relaciones y de los tipos de relaciones: alta, eliminación y modificación.
- Consultar el glosario. La búsqueda de un término dentro del glosario se puede realizar por su letra inicial o por el término completo, y según la lengua a la que pertenece.

Es interesante destacar como novedad frente a otras herramientas similares que, en los resultados mostrados al usuario en cada momento aparecen elementos que sirven de enlace para obtener información sobre ellos. Por ejemplo, los términos relacionados con

uno dado aparecen resaltados y si se pincha sobre ellos, se puede acceder a su ficha dentro del glosario.

### Ejemplos de uso de la herramienta Calíope

Como se ha podido deducir, Calíope no es inicialmente una herramienta de enseñanza en un sentido estricto –no explica conceptos gramaticales, no propone ejercicios, no corrige al alumno,–. Sin embargo, su uso puede servir como apoyo tanto a alumnos como a profesores. Mediante los resultados de las consultas a la aplicación se pueden inferir reglas gramaticales o normas de uso, fundamentales para el aprendizaje de la redacción técnica. Y viceversa, dada una regla común como punto de partida, se puede aplicar para buscar ejemplos relevantes o representativos en el corpus. Veamos algunas muestras de las posibilidades que proporciona el sistema, siempre que se trabaje con un corpus adecuado y significativo.

a) Se puede averiguar qué términos o qué variantes terminológicas son más habituales dentro de un dominio o subdominio. Por ejemplo, la frecuencia de las palabras puede indicar que entre dos sinónimos, uno apenas es utilizado (como ocurre, en el ámbito informático con los términos *evaluation* y *assessment*) y, además, gracias a las concordancias, puede observarse en qué casos concretos se emplea cada uno (véanse FIGURA 1 y FIGURA 2). En los resultados de las concordancias se muestra el término en la columna central y recuadrado en verde y acompañado a izquierda y derecha del nº de palabras especificado como parámetro. Pinchando sobre el término se accede al texto completo en el punto correspondiente a la ocurrencia de la que se trate.

Calíope Corpus textual y Glosario de términos informáticos											
<a href="#">Salir (abandonar la sesión)</a> <a href="#">Administrar usuarios</a> <a href="#">Administrar textos</a> <a href="#">Consultar Corpus</a> <a href="#">Consultar Glosario</a>											
Operaciones: Operación: Concordancia Palabra: evaluation Entorno: 3 Textos seleccionados: Todos Idioma: Español, Inglés Fecha: Lunes 21-Mayo-2007, 09:36:00 • Calíope											
Concordancias											
[13]	also	be	used	as	an	evaluation	metric	to	assess	the	price
[13]	Analysed	and	their	Special	Features	Evaluation	Criteria	This	analysis	was	carried
[13]	and	other	related	standards	TMS	evaluation	criteria	the	electronic	interchange	of
[13]	MARTIF	ISO	DIS	12200	TMS	Evaluation	Criteria	The	identification	of	the
[13]	participating	academic	organisations	will	devise	evaluation	criteria	that	will	help	in
[13]	of	the	results	of	the	evaluation	will	recommend	reading	lists	attendance
[13]	methodologies	including	terminology	interchange	comparative	evaluation	of	terminology	management	systems	and
[25]	and	to	participate	in	prototype	evaluations	or	beta	testing	Reconcile	conflicting
[25]	we	generally	use	structured	prototype	evaluation	scripts	intended	to	generate	answers
[25]	we	have	explored	with	prototype	evaluation	scripts	Does	this	program	do
[25]	verbal	responses	to	my	prototype	evaluation	questions	From	Requirements	to	Code
[28]	machines	such	as	the	various	evaluation	boards	it	is	possible	to
[30]	with	your	after	the	fact	evaluation	of	how	important	the	requirement
[41]	string	Martin	Geisler	after	the	evaluation	without	any	sign	of	all
TOTAL OCURRENCIAS ENCONTRADAS: 14											

FIGURA 1: Resultados de concordancias para *evaluation*

Calíope Corpus textual y Glosario de términos informáticos

Salir (abandonar la sesión)

Administrar usuarios Administrar textos Consultar Corpus Consultar Glosario

Concordancia

Operación: Concordancia Palabra: assessment Entorno: 5 Textos seleccionados: Todos Idioma: Español, Inglés Fecha: Lunes, 21-Mayo-2007, 09:28:00

[3]	such	a	manner	that	unambiguous	assessment	criteria	pass	fail	or	some
[3]	are	subjective	and	a	conclusive	assessment	of	quality	requires	a	technical
[10]	new	alliances	have	spawned	conflicting	assessments	of	how	customers	should	move
[15]	to	do	so	Respect	developers	assessments	of	cost	and	feasibility	Set
[15]	To	respect	a	developer	s	assessment	of	cost	and	feasibility	All

TOTAL OCURRENCIAS ENCONTRADAS: 5

FIGURA 2: Resultados de concordancias para *assessment*

b) El hecho de mantener información en distintas lenguas, permite estudiar si las características de una lengua se cumplen también en la otra. Por ejemplo, el neologismo inglés *download* es, por ahora, exclusivo del ámbito de Internet, mientras que sus equivalentes en español, “descargar(se)” o “bajar(se)”, tiene un uso mucho más general. Se ha perdido en este caso la especificidad del término inglés y han aparecido nuevos usos fraseológicos en español como “descargar una película, una canción, etc.” Esto se puede comprobar examinando el tipo, el campo y la fuente de los textos en los que aparece el término en concreto (véase FIGURA 3).

Calíope Corpus textual y Glosario de términos informáticos

Salir (abandonar la sesión)

Administrar usuarios Administrar textos Consultar Corpus Consultar Glosario

Modificación de Campo de Texto

Signatura: 1

Campo: Ingeniería del software

Modificación de Tipo de Texto

Signatura: 1

Tipo: Texto científico-técnico

Especificador:

Modificación de Fuente de Texto

Tipo: Texto electrónico

Signatura: 1

Título: Frases aelfe

Autor:

Lugar/URL:

Editorial:

Fecha: día: 19 año: 04 2007

FIGURA 3: Ejemplo de campo, tipo y fuente asociados a un texto del corpus

c) Las concordancias también ayudan a establecer los participantes en las construcciones sintáctico-semánticas de las lenguas. Por ejemplo, la observación de las concordancias de *uninstall*, lleva a la conclusión de que muy frecuentemente va acompañada de la palabra *program(s)*. Esta situación se puede almacenar en el glosario de términos mediante una relación (véanse FIGURA 4 y parte inferior de FIGURA 5). Las relaciones siempre se definen entre dos términos del glosario y, en los casos en que sea necesario, incluyen la preposición que ha de ir entre ellos. Está previsto además, la

posibilidad de definir nuevos tipos de relaciones cuando no se disponga de uno adecuado.



FIGURA 4: Establecimiento de una relación entre *uninstall* y *program*

d) Otra ventaja que también ofrece la aplicación es el hecho de que pinchando en *program* se puede acceder a su ficha en el glosario y ver con qué otros verbos/términos está relacionado (véase FIGURA 5) y estudiar si existe alguna conexión entre ellos. Los contextos del término se muestran en un listado, encuadrados con fondo gris y borde punteado. El término de la ficha aparece subrayado para localizarlo fácilmente dentro de su contexto. A la izquierda del cuadro se encuentra, entre corchetes, el identificador del texto del cual se ha extraído el contexto.

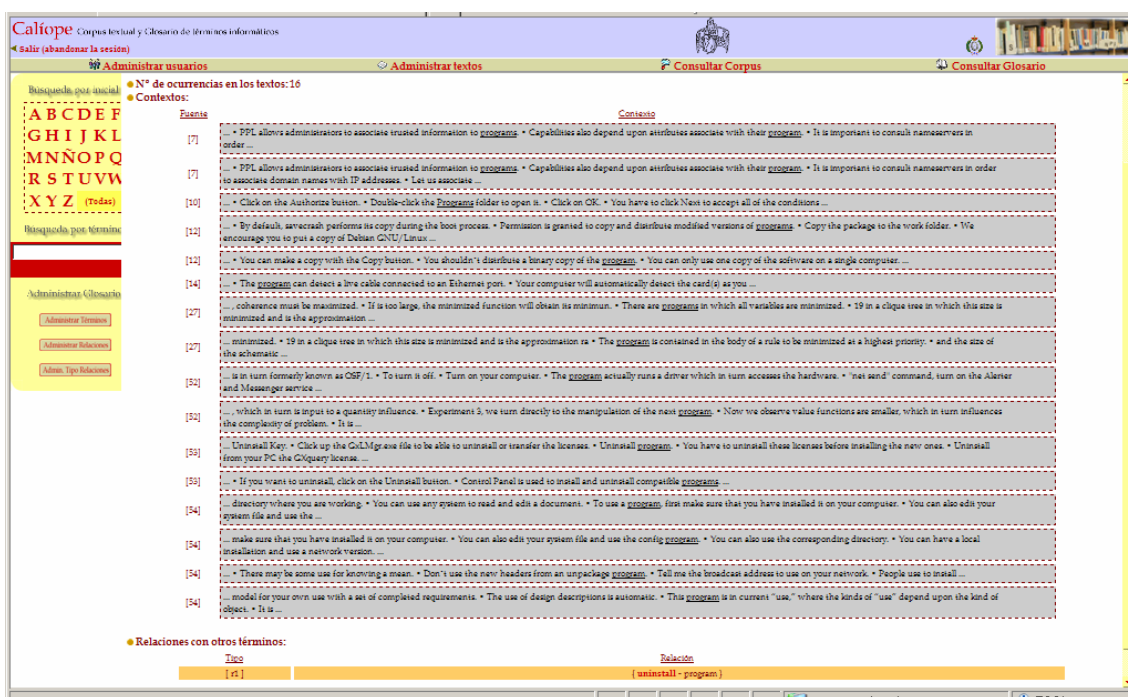


FIGURA 5: Información guardada en el glosario relativa al contexto y las relaciones de *program*

Además de estos casos de uso directos, conviene apuntar otros casos que podrían denominarse indirectos, empleando los resultados de las consultas y el contenido del glosario para la elaboración de ejercicios. Por ejemplo, mediante textos con huecos que se han de rellenar con los términos adecuados, o conjuntos de términos que se han de ordenar convenientemente, etc.

Por último, también es importante señalar que Calíope podría utilizarse tanto en clases presenciales (con el equipamiento preciso), como integrarse en el currículum de un curso a distancia por Internet ya que está proyectado para estar disponible en línea.

## Conclusiones

En este trabajo hemos presentado Calíope, una aplicación informática que permite gestionar y hacer consultas sobre un corpus y un glosario de términos que aparecen en él y hemos comprobado su utilidad en la enseñanza de LSP (en español o inglés) dentro del ámbito informático. Además de mostrar en qué contextos se utiliza cada término, la aplicación permite establecer (y ver) relaciones (semánticas) entre términos, posibilitando así tener contenido elaborado, no simplemente datos sin analizar, aspecto que la diferencia de otras herramientas que trabajan con corpus y glosarios. Asimismo cada elemento está interconectado con el resto de elementos del sistema de forma que se puede ir navegando por ellos de manera sencilla para el usuario. Todas estas características hacen que su uso resulte también muy atractivo en el ámbito de la traducción y la redacción (de textos informáticos) en español o inglés.

Como casi en cualquier proyecto informático, la herramienta que se ha descrito no constituye un sistema que se dé por concluido. Evidentemente, la aplicación admite mejoras de las funcionalidades existentes y también futuras ampliaciones. Entre las inmediatas previstas podemos destacar las siguientes: aumentar las opciones en la búsqueda de concordancias (por ejemplo, utilizando términos compuestos o caracteres comodín, empleando variantes morfológicas, tiempo verbal admitiendo la búsqueda de coocurrencias, etc.); posibilitar que el listado de términos esté ordenado según criterios distintos al alfabético (por ejemplo, según número de ocurrencias); permitir cualquier formato en los textos; incorporar el tratamiento de textos anotados y facilitar su anotación; y mejorar la administración de los usuarios (especialmente en el alta mediante archivo).

Finalmente, es importante indicar que aunque la aplicación está pensada para trabajar con textos y términos en inglés o en español dentro del ámbito informático, su utilización es fácilmente extensible a otros ámbitos y el diseño y las características de la herramienta hacen muy sencilla la incorporación de otras lenguas.

## Referencias

- Aguado de Cea, G. (2007). "La fraseología en las lenguas de especialidad" en Alcaraz, E. Mateo, J. y Yus, F. (eds.), *Las lenguas profesionales y académicas*, Madrid: Ariel
- Barahona, R. (2006). *Calíope: Corpus textual y Glosario de Términos Informáticos*. Trabajo Fin de Carrera. Facultad de Informática, UPM.
- Bowker, L. & Pearson, J. (2002). *Working with specialized language. A practical guide to using corpora*. London: Routledge.
- Cabré, M.T.(2003). "Theories of terminology". *Terminology*, 9,2: 163-169.
- L'Homme, M.C. (2000): "Understanding Specialized Lexical Combinations". *Terminology*, 6,1: 89-110.
- Mel'cuk, I. A. (1988). *Dependency syntax: Theory and practice*. Albany: State University of New York Press.

Sánchez, A. (2000). "Language Teaching before and after 'Digitalized Corpora'. Three main issues" en Cantos, P. & Sánchez (eds.), A. *Corpus-based research in English Language and Linguistics*, 9,1. Murcia: Cuadernos de Filología Inglesa.

Temmerman, R. (2000). *Towards new ways of Terminology Description*. Amsterdam/Philadelphia: John Benjamins.